WHAT IS CLAIMED IS:

1. A computer-based system for creating a targeted collection of sequences from a dataset comprising sequence identifiers corresponding to natural complex biopolymer sequences and linked to corresponding annotations, the system comprising:

a) a search function which searches the annotations of the dataset according to a user-defined criterion and outputs a first subset of the dataset restricted by the criterion;

b) a redundancy reducing function which compares the first subset with a first database correlating the sequence identifiers of the first subset with syngeneic biopolymers and outputs a second subset of the dataset having reduced unique, natural complex biopolymer redundancy relative to the first subset;

c) a selection function which applies to the second subset a user-defined selection parameter and outputs a third subset restricted relative to the second subset by the parameter; and

d) a tabulation function which creates and outputs the targeted collection of sequences in the form of a data table comprising, configurable by and sortable by the sequence identifiers of the third subset.

2. A system according to claim 1, wherein the criterion is selected from the group consisting of a keyword and a concept.

3. A system according to claim 1, wherein the criterion is one of a plurality of user-defined criteria, and the search function searches the annotations of the dataset according to the criteria and outputs a first subset of the dataset restricted by the criteria.

4. A system according to claim 1, wherein the criterion is one of a plurality of user-defined criteria, and the search function searches the annotations of the dataset according to the criteria and outputs a first subset of the dataset restricted by the criteria, wherein the criteria include multiple keywords.

5. A system according to claim 1, wherein the dataset is selected from the group consisting of GenBank, Medline and KEGG.

6. A system according to claim 1, wherein the dataset is one of a plurality of datasets, and the search function searches the annotations of the datasets according to the user-defined criterion and outputs a first subset of the datasets restricted by the criterion.

7. A system according to claim 1, wherein the database is selected from the group consisting of UniGene and LocusLink.

8. A system according to claim 1, wherein the database is one of a plurality of databases correlating the sequence identifiers of the first subset with syngeneic biopolymers, and the redundancy reducing function compares the first subset with the databases and outputs the second subset of the dataset.

9. A system according to claim 1, wherein the parameter is selected from the group consisting of source, species, author and pathway.

10. A system according to claim 1, wherein the parameter is one of a plurality of user-defined selection parameters, and the selection function applies to the second subset the parameters and outputs the third subset restricted relative to the second subset by the parameters.

11. A system according to claim 1, wherein the redundancy reducing function outputs a second subset of the dataset which eliminates unique, natural complex biopolymer redundancy relative to the first subset.

12. A system according to claim 1, further comprising an expansion function which searches a second database for synonyms of the sequence identifiers of the first, second or third subset.

13. A computer-based method for creating a targeted collection of sequences from a dataset comprising sequence identifiers corresponding to natural complex biopolymer sequences and linked to corresponding annotations, the method comprising computer-implemented steps of:

a) searching with a computer the annotations of the dataset according to a user-defined

criterion and outputting a first subset of the dataset restricted by the criterion;

b) comparing with the computer the first subset with a database correlating the sequence identifiers of the first subset with syngeneic biopolymers and outputting a second subset of the dataset having reduced unique, natural complex biopolymer redundancy relative to the first subset;

c) applying to the second subset a user-defined selection parameter and outputting a third subset restricted relative to the second subset by the parameter; and

d) creating and outputting the targeted collection of sequences in the form of a data table comprising, configurable by and sortable by the sequence identifiers of the third subset

14. A computer-based system for creating a targeted collection of sequences from a plurality of datasets comprising sequence identifiers corresponding to natural complex biopolymer sequences, the system comprising:

a) a merge and redundancy reducing function which compares the datasets with a database correlating the sequence identifiers with syngeneic biopolymers and creates a subset of the sum of the datasets having reduced unique, natural complex biopolymer redundancy relative to the sum; and

b) a tabulation function which creates and outputs the targeted collection of sequences in the form of a data table comprising, configurable by and sortable by the sequence identifiers of the subset.

15. A system according to claim 14, wherein the merge and redundancy reducing function further comprises a selection function which applies a user-defined selection parameter whereby the subset is restricted relative to the sum of the datasets by the parameter.

16. A system according to claim 14, wherein the merge and redundancy reducing function further comprises a selection function which applies a user-defined selection parameter whereby the subset is restricted relative to the sum of the datasets by the parameter, wherein the parameter is selected from the group consisting of source, author and pathway.

17. A computer-based method for creating a targeted collection of sequences from a plurality

of datasets comprising sequence identifiers corresponding to natural complex biopolymer sequences, the method comprising computer-implemented steps of:

a) comparing the datasets with a database correlating the sequence identifiers with syngeneic biopolymers and creating a subset of the sum of the datasets having reduced unique, natural complex biopolymer redundancy relative to the sum; and

b) creating and outputting the targeted collection of sequences in the form of a data table comprising, configurable by and sortable by the sequence identifiers of the subset.

18. A computer-based system for creating a targeted collection of sequences from a dataset comprising sequence identifiers corresponding to natural complex biopolymer sequences and linked to corresponding first annotations, the system comprising:

a) an integration function which merges the dataset with a database comprising second annotations attributable to and correlated with at least a subset of the sequence identifiers or sequences of the dataset and which links the second annotations to the corresponding sequence identifiers of the subset; and

b) a tabulation function which creates and outputs the targeted collection of sequences in the form of a data table comprising, configurable by and sortable by the sequence identifiers of the subset and the second annotations.

19. A system according to claim 18, wherein the second annotations comprise data attributable to and correlated with at least a subset of the sequence identifiers or sequences of the dataset, said data selected from the group consisting of: gene expression data, sequencing data, genotype data, polymorphism data and clinical data.

20. A computer-based method for creating a targeted collection of sequences from a dataset comprising sequence identifiers corresponding to natural complex biopolymer sequences and linked to corresponding first annotations, the method comprising computer-implemented steps of:

a) merging the dataset with a database comprising second annotations attributable to and correlated with at least a subset of the sequence identifiers or sequences of the dataset and linking the second annotations to the corresponding sequence identifiers of the subset; and

b) creating and outputting the targeted collection of sequences in the form of a data table comprising, configurable by and sortable by the sequence identifiers of the subset and the second annotations.

21. A system according to claim 1, further comprising:

a second computer-based system for creating a targeted collection of sequences from a plurality of datasets comprising sequence identifiers corresponding to natural complex biopolymer sequences, the second system comprising:

a) a merge and redundancy reducing function which compares the datasets with a database correlating the sequence identifiers with syngeneic biopolymers and creates a subset of the sum of the datasets having reduced unique, natural complex biopolymer redundancy relative to the sum; and

b) a tabulation function which creates and outputs the targeted collection of sequences in the form of a data table comprising, configurable by and sortable by the sequence identifiers of the subset.

22. A system according to claim 1, further comprising:

a second computer-based system for creating a targeted collection of sequences from a dataset comprising sequence identifiers corresponding to natural complex biopolymer sequences and linked to corresponding first annotations, the second system comprising:

a) an integration function which merges the dataset with a database comprising second annotations attributable to and correlated with at least a subset of the sequence identifiers or sequences of the dataset and which links the second annotations to the corresponding sequence identifiers of the subset; and

b) a tabulation function which creates and outputs the targeted collection of sequences in the form of a data table comprising, configurable by and sortable by the sequence identifiers of the subset and the second annotations.

23. A system according to claim 1, further comprising:

a second computer-based system for creating a targeted collection of sequences from a plurality of datasets comprising sequence identifiers corresponding to natural complex

biopolymer sequences, the second system comprising:

a) a merge and redundancy reducing function which compares the datasets with a database correlating the sequence identifiers with syngeneic biopolymers and creates a subset of the sum of the datasets having reduced unique, natural complex biopolymer redundancy relative to the sum; and

b) a tabulation function which creates and outputs the targeted collection of sequences in the form of a data table comprising, configurable by and sortable by the sequence identifiers of the subset; and,

a third computer-based system for creating a targeted collection of sequences from a dataset comprising sequence identifiers corresponding to natural complex biopolymer sequences and linked to corresponding first annotations, the third system comprising:

a) an integration function which merges the dataset with a database comprising second annotations attributable to and correlated with at least a subset of the sequence identifiers or sequences of the dataset and which links the second annotations to the corresponding sequence identifiers of the subset; and

b) a tabulation function which creates and outputs the targeted collection of sequences in the form of a data table comprising, configurable by and sortable by the sequence identifiers of the subset and the second annotations.

24. A system according to claim 1, wherein the system is ARROGANT.